

用于自然植物细粒度分类的目标定位和区域选择网络

黄文龙¹ 马祥星²

HUANG Wenlong MA Xiangxing

摘要

近年来,植物中的细粒度分类问题已经成为计算机视觉领域热门的方向之一。现有的大部分细粒度分类方法只专注于如何寻找更有区分性的部位,忽视了目标的整体结构有助于网络的分类和区分性部位的定位。基于注意力定位关键区域的方法大多都是利用局部注意力来直接定位的,这些方法无法判断所定位到的区域在全局所有区域中是否是最有区分性的。为了解决以上问题,提出一个模板定位和区域选择网络从粗粒度到细粒度渐进式的学习目标的特征。具体来说,目标定位模块可以定位到完整的目标,排除背景的干扰,强化网络提取目标整体结构特征的能力,有助于准确分类和后续区分性区域的定位。区域选择模块则通过全局相关性来衡量数千个区域在全局上的重要程度,根据重要性来选择最有区分性的区域,使网络学习目标的细粒度特征。在 iNaturalist-2021Mini 和 iNaturalist-2018 数据集上的实验表明,所提出的方法可以达到优异的效果。

关键词

细粒度分类;植物分类;注意力机制;卷积神经网络

doi: 10.3969/j.issn.1672-9528.2024.03.025

0 引言

植物在生态保护、气候调节和经济发展方面扮演着关键角色。它们对于维护生态平衡、减轻温室效应和气候变化,以及推动可持续的经济繁荣起到了至关重要的作用。不同植物在整体外观结构上具有相似性,很多情况下只在细微处才有区别,而同一种植物也可能因为不同的视角表现出较大的差异性,这些都是识别植物所面临的问题。目前,植物的分类主要依靠领域专家的人工识别,导致基于图像的入侵植物识别任务工作量巨大且效率不高。目前,细粒度视觉分类(FGVC)已经是计算机视觉领域最热门的方向之一,近些年出现了大量优秀的方法。细粒度分类旨在对属于同一超类的子类别进行分类,例如植物(花、树)的各种子类。基于细粒度分类的植物识别可以有效解决以上问题。

由于细粒度分类存在的类间相似性和类内差异性问题,现有的算法主要是通过注意力或者检测等方法找到区分性区域来识别不同的细粒度类别^[1-2]。尽管这种方法已经取得了很大进展,但是目前还存在一些问题。(1)大量的研究工作只致力于研究如何更好地捕捉局部区域中的区别性部位,这些工作忽视了目标整体结构的学习对于提升网络性能的研究。

而且准确地定位判别性区域也严重依赖目标的整体结构。(2)已有的各种基于注意力定位关键区域的方法使用的是通道或者空间注意力等局部注意力。他们没有研究自注意力所建立的全局相似关系对于定位关键区域的作用,通过这种全局相关性可以衡量网络生成的数千个注意力区域的重要性程度,从中选择最重要的区域。

为了解决以上两个问题,本文提出了一个目标增强和区域选择网络(object enhancement and region selection network, OERS-Net)来模拟人类识别植物的过程。其中,使用目标增强和定位模块来定位目标,以方便网络更好地学习目标的整体结构信息,判断目标的大致类别。在定位到目标的基础上,区域选择模块从数千个注意力区域中寻找最有区分性的区域,进一步学习目标判别性区域的特征,进行细粒度类别的分辨。

本文的主要贡献总结如下。

(1)提出了一个用于定位目标的目标增强和定位模块(object enhancement and location attention, OELA),它可以准确定位到目标的位置,使网络更好地学习目标的全局结构特征,提升网络性能,同时排除背景的干扰,有利于下一步进行区域的定位。

(2)提出了一个用于定位多个区分性区域的区域选择模块(multi-region selection attention, MRSA),它利用自注意力的全局性关系选择全局结构中重要的多个部位,使网

1. 三峡大学计算机与信息学院 湖北宜昌 443000

2. 皖西学院金融与数学学院 安徽六安 237000

络有效学习细粒度级的区分性特征。

(3) 在两个数据集 (iNAT21mini-plant, iNAT18-plant) 上进行了实验, 证明了本文网络的有效性。

1 相关工作

过去这些年, 细粒度分类方向出现了大量的方法来解决该领域特有的类间相似性和类内差异性问题, 其中主流的方法是通过寻找关键的判别性区域进行分类。

1.1 定位关键区域

通过定位关键区域的方法, 可以采用注意力或者目标检测等方法来定位区分性部位的信息并进行分类。TASN^[3] 使用三线注意力采样网络, 生成结构采样图像和细节采样图像, 网络在学习结构采样图片信息的同时通过蒸馏的方法强化网络对细节信息的学习。与现有的直接定位关键区域不同, CAL 提出了反事实注意力, 通过生成虚假的注意力区域来施加反事实干预, 直接有效地监督网络学习到的注意力区域是不是具有区分性, 而不是通过结果来间接监督注意力区域的学习。

然而, 上述方法只专注于如何寻找更有区分性的区域来提升网络性能。与这些工作不同的是, 本文认为目标的整体结构特征和区分性区域的特征一样重要。在研究网络如何寻找有区分性的区域同时, 同样看重网络学习目标整体结构特征的能力。学习目标的结构特征不仅能提升网络识别的性能, 还有利于网络寻找区分性区域。

1.2 通过注意力的方法

细粒度分类中的注意力在进行关键区域定位的同时, 还有一项工作是使用注意力来强化特征图, 然后把强化后的特征图送入网络进行分类。PCA-Net^[4] 网络一次性输入两张同类别图片, 通过同类图片之间通道的相似性来学习共同的区域特征。和现有直接提出注意力模块的方法不同, PMG^[5] 通过使用随机拼图生成器把图片打乱成不同大小的 patch, 这种做法破坏了图像的整体结构信息, 保留了细节部位, 从而使网络只能关注局部区域, 最后送入完整的图片。

如图 1 所示, 图像分支学习原始图像的信息。OELA 通过图像分支的特征图定位得到目标图片。目标分支学习目标的整体结构信息。MRSA 根据目标分支的特征图选择多个区分性区域。区域分支学习目标的多细节区域的信息。以上工作中的各种空间或通道注意力大都是基于卷积的局部相关性生成的。得益于自注意力优秀的长距离依赖, 本文的多区域选择 (MRSA) 模块利用了这种全局相关性, 通过全局关系矩阵来衡量不同的注意力区域在全局所有区域的重要性程度, 最后选择重要的那些区域供网络学习。

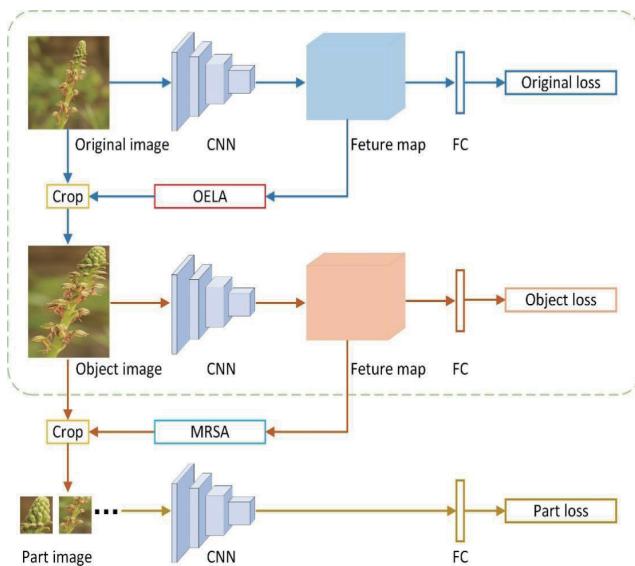


图 1 OERS-Net 的网络结构图

2 方法

2.1 目标增强和多区域选择网络 (OERS-Net)

本文提出的 OERS 网络包含三个部分, 即用于提取特征图的卷积神经网络、定位目标结构的 OELA 模块、关注多个判别性区域的 MRSA 模块。为了使网络充分学习 OELA 模块所定位的整体结构特征以及 MRSA 模块关注的多个细节区域特征, 本文提出了一个具有三支的网络来分别学习图像信息、目标结构信息以及多个判别性区域信息。

本文的网络结构如图 1 所示。在训练阶段, 当图像输入网络后, 图像网络学习目标和背景的关系, 目标网络学习 OELA 模块所定位的目标结构信息, 区域网络进一步学习 MRSA 所选择的多个细粒度级信息区域。在推理阶段, 本文的网络使用目标分支的结果作为预测结果。部件网络只用于训练, 不用于推理。三个网络共享卷积层和全连接层的参数, 这样不仅节约了模型参数, 还能使该网络在提取目标完整结构信息的基础上更侧重于关注区分性区域的信息。

先根据特征图得到多个不同大小的激活区域, 然后使用腐蚀和膨胀把包含目标信息的多个区域连通, 得到完整的目标包围框, 最后定位到原图裁剪并放大。

2.2 目标增强和定位模块 (OELA)

细粒度分类网络的性能在很大程度上依赖于能否找到准确的判别性部位, 学习关键区域的特征固然能提升模型分类的准确性, 但是目标整体结构特征同样重要, 仅依靠关键区域的定位限制了细粒度分类网络性能的进一步提升。为了解决该问题, 本文提出了一个目标增强和定位模块 (OELA) 来寻找目标区域, 强化网络学习整体结构特征的能力。OELA 模块根据特征图来生成目标包围框的位置坐标, 映射到原图进行定位, 从而可以排除背景的干扰, 使网络学习前景目标的整体结构特征, 提升网络分类的准确性。除此之外,

由于排除了背景等无关因素的干扰，在此基础上进行关键区域定位，可以使网络学习到更细节的特征。本文的 OELA 模块如图 2 所示。

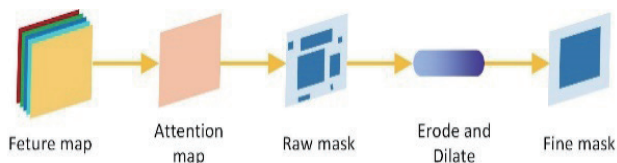


图 2 OELA 模块的流程图

2.2.1 生成激活区域图

本文提出的 OELA 模块也是通过空间注意力的方式定位目标。给定一张图片 X ，把它送入卷积神经网络提取特征，取网络最后一层的特征图 $F \in \mathbb{R}^{C \times W \times H}$ ，用 C 来描述特征图的通道， $H \times W$ 是特征图的高和宽。特征图 F 在通道维度相加得到激活图 $A \in \mathbb{R}^{1 \times W \times H}$ ，其公式为：

$$A = \sum_{i=0}^{C-1} F_i \quad (1)$$

F_i 代表特征图第 C 个通道的特征， C 是通道数。试图通过 A 寻找目标所在哪些区域。具体做法是先把激活图 A 取均值得到 z ，其公式为：

$$z = \frac{\sum_{x=0}^{W-1} \sum_{y=0}^{H-1} A(x, y)}{H \times W} \quad (2)$$

激活图 A 中大于 z 的点属于目标的一部分，对应位置赋值为 1，小于 z 的点是和目标无关的背景区域，赋值为 0，其公式为：

$$\tilde{A}_{(x,y)} = \begin{cases} 1 & \text{if } A_{(x,y)} > z \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

式中： (x, y) 代表特征图上特征点的坐标，由此得到了一个包含多个激活区域的激活图 $\tilde{A}$ 。

2.2.2 得到覆盖目标区域的掩码

得到的注意力图 $\tilde{A}$ 中存在多个大小不一的连通区域，越大的连通区域包含的目标信息越多。但在很多情况下，即使是最大的连通区域，也没有完整地覆盖住目标区域，仅仅只是包含了目标较大的一部分结构信息。因此，通过引入腐蚀和膨胀算法，把包含目标信息的不同激活区域连通，以形成完整的目标区域。由于真实的目标是连续完整的结构，所以包含目标不同部分信息的连通区域也是接近的，只需把包含目标信息的相近区域连通即可，最后选择连通区域最大的地方形成包围框，由此得到了包含完整目标区域的掩码。最后裁剪并放大，得到目标图片。

2.3 区域选择模块 (MRSA)

现有的各种基于注意力寻找关键区域的方法已经被证明在细粒度分类中是有效的。目前已有的各种通道或者空间注意力都是利用局部相关性来定位的，这些方法并没有探究全

局相关性对于定位关键区域的作用。因此，本文提出了一个多区域选择自注意力 (MRSA) 模块，如图 3 所示。

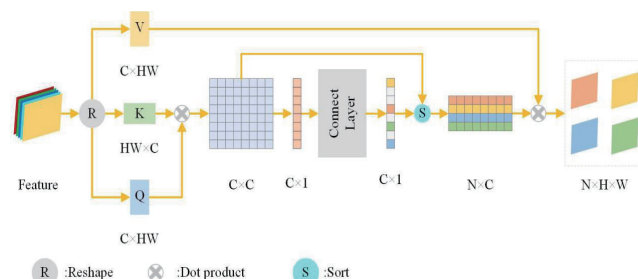


图 3 MRSA 模块的流程图

先通过自注意力的方式得到通道间关系矩阵 M ， M 的每一行按照重要性进行排序，选择最有区分性的区域作为部分注意力图。

MRSA 模块通过自注意力来衡量网络每个注意力区域在全局上所有区域的重要性。根据已有的经验，每个通道的卷积特征对应网络的一个视觉区域，所有这些通道共同组成了完整的图像特征。为了使 MRSA 模块能够选择其中最重要的一些部分，利用自注意力的全局相关性为每个通道赋予重要性值，然后通过重要性值，可以得到每个通道在全局所有通道中的重要性程度，最后进行排序，从而选择重要性较高的特征通道。

2.3.1 生成全局关系矩阵

MRSA 将目标网络提取的目标特征图 $F \in \mathbb{R}^{C \times W \times H}$ 作为输入， C 、 H 和 W 分别代表通道、高和宽。首先把特征图 F 重塑为 $Q \in \mathbb{R}^{C \times HW}$ ， $K \in \mathbb{R}^{HW \times C}$ ， $V \in \mathbb{R}^{C \times HW}$ ，把 Q 和 K 相乘得到通道间的关系矩阵 $M_{qk} \in \mathbb{R}^{C \times C}$ ，然后通过 softmax 函数将矩阵的每一行带权归一化，从而得到通道上的全局注意力矩阵 $M \in \mathbb{R}^{C \times C}$ ，计算公式为：

$$M = \text{Softmax}(QK) \quad (4)$$

2.3.2 衡量每个通道在全局上的重要性

把 M 通过求平均并送入全连接层来为每个通道生成一个重要性指标，以筛选哪些通道包含的区域特征对细粒度分类最有用。当它接近 0 时，代表该通道与所关注的特征中心距离较远，这些通道关注的是无效区域，对于结果的帮助有限。为了集中特征中心附近的通道特征区域，根据重要性值把 M 的行按照降序进行排列，选择前 N 行作为最终的部件注意力 $A_n \in \mathbb{R}^{N \times C}$ ，把 A_n 乘以 V 得到区域特征图 $F_n \in \mathbb{R}^{N \times HW}$ ， N 为区域特征图的通道数，即包含了 N 个区分性区域。

通过 MRSA 模块，得到了 N 个显著的区分性区域。最后将这 N 个部分所包含的位置信息映射到原图，裁剪得到 N 个细节区域，送入区域网络以学习细节区域信息。

2.4 损失

本文网络的三个分支均使用交叉熵损失作为分类损失，

因此总体损失函数为:

$$L(\hat{Y}_{image}, \hat{Y}_{object}, \hat{Y}_{part}) = -\log(\hat{Y}_{image}) - \log(\hat{Y}_{object}) - \sum_{n=0}^{N-1} \log(\hat{Y}_{part}) \quad (5)$$

式中: $\hat{Y}_{image}$ 是图像分支对于图片的分类结果, $\hat{Y}_{object}$ 是目标分支对于截取的目标图片的分类结果, $\hat{Y}_{part}$ 是部件分支对于所有区分性区域的分类结果, 总损失是三个分支的损失之和, 一起对网络进行优化。

至此, 本文的网络可以由粗到细, 逐步学习图像级到目标级再到细粒度级的信息。

3 实验

本节首先介绍训练的设置, 包括数据集和训练方法以及所对比的网络。然后, 在三个数据集, 包括 18-Plant、21-Plant 和 FGVC-Aircraft, 把本文的方法和最先进的细粒度分类方法进行比较。

3.1 数据集

为了研究所提出的网络的性能, 在两个细粒度分类的数据集上进行了实验和评估, 包括 21-Plant 和 18-Plant, 如表 1 所示。21-plant 和 18-plant 的数据取自 iNaturalist-2021Mini^[6] 和 iNaturalist-2018^[7] 的植物大类, 以方便评估本文的网络在植物分类上的性能。

表 1 FGVC-Aircraft、21-Plant、18-Plant 数据集介绍

Dataset	Categories	Training Set	Testing Set
21-Plant	4721	213 550	42 710
18-Plant	2917	118 800	8751

3.2 消融实验

为了验证本文提出的 OELA 和 MRSA 模块的有效性, 在所使用的三个数据集上均做了消融实验, 除了所提出的模块, 数据增强方式和学习率等超参数设置均保持一致, Baseline 使用的是 ResNet 50。从实验结果来看, 通过 OELA 模块定位的目标来训练网络比 Baseline 拥有更好的效果, 证明了目标的整体结构确实有利于提升网络的性能。MRSA 所关注的多个区分性区域对于网络性能的提升同样明显。当同时使用 OELA 和 MRSA 模块时, 网络达到了最好的效果。实验结果如表 2 所示。

表 2 在 iNaturalist-21 和 iNaturalist-18 的植物大类上的消融实验

Model	21-Plant	18-Plant
Baseline	70.5	66.5
OELA	74.7	70.8
MRSA	75.5	72.4
OERS-Net	76.8	73.7

3.3 对比试验

在所用的两个数据集上与以下方法进行比较, 因为它们具有最先进的性能。为了排除公平的比较, 探究本文方法的有效性, 选择使用卷积网络作为 backbone 的网络进行比较, 没有使用人工标记的部分区域数据。根据结果可知, 在两个数据集上均达到了最好的效果。在 21mini-plant 和 18-plant 的对比实验如表 3。

表 3 在 iNaturalist-21 和 iNaturalist-18 的植物大类上的对比实验

Model	21-Plant	18-Plant
TASN (19-CVPR)	72.7	69.3
PMG (20-ECCV) [8]	75.3	67.5
PCA-Net (21-VCIP)	73.0	70.6
MMAL (21-MMM)	76.7	73.4
P2P-NET (22-CVPR) [9]	74.8	71.9
MHEM (22-TNNLS) [10]	73.1	70.4
Ours(OERS-Net)	76.8	73.7

4 结论

本文提出了用于定位目标完整结构的 OELA 模块和筛选并定位多个重要区域的 MRSA 模块。这两个模块均没有引入额外的参数, 不需要进行训练。本文的方法在多个数据集上均具有良好的结果。后续的工作是研究如何把包含相同部位的不同通道的特征图进行融合。

参考文献:

- [1]RAO Y, CHEN G, LU J, et al. Counterfactual attention learning for fine-grained visual categorization and re-identification[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision.Piscataway:IEEE,2021:1025-1034.
- [2]ZHANG F, LI M, ZHAI G, et al.Multi-branch and multi-scale attention learning for fine-grained visual categorization[C]// International Conference on Multi-Media Modeling. Cham:Springer Nature Switzerland AG,2021:136-147.
- [3]ZHENG H, FU J, ZHA Z J, et al.Looking for the devil in the details: learning trilinear attention sampling network for fine-grained image recognition[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 5012-5021.
- [4]ZHANG T, CHANG D, MA Z, et al.Progressive co-attention network for fine-grained visual classification[C]// 2021 International Conference on Visual Communications and Image Processing (VCIP). Piscataway: IEEE, 2021:1-5.
- [5]DU R, CHANG D, BHUNIA A K, et al. Fine-grained visual classification via progressive multi-granularity training of jigsaw patches[C]//European Conference on Computer Vision. Cham: Springer International Publishing, 2020:153-168.

基于 TUSB2046 的多处理器调试模块设计与实现

吴 超¹ 汪利建¹ 王洋洋¹
WU Chao WANG Lijian WANG Yangyang

摘 要

针对航空电子设备调试过程中,传统的模块调试方法需要多次在机柜中插拔模块,易造成模块损坏,且调试效率低下的问题,基于 USB_HUB 控制芯片 TUSB2046IBVF 设计了一款同时适用于不超过 10 模块的实时在线调试电路并实现为调试模块。首先介绍了设计方案的物理架构,其次介绍了模块核心芯片的供电方式、时钟电路、负载的过流保护、与机柜的连接器定义等,最后通过对模块的测试,达到了预定的功能,效果良好。

关键词

通用串行总线集线器 (USB_HUB); 调试模块; 多处理器; 联合测试行为组织 (JTAG)

doi: 10.3969/j.issn.1672-9528.2024.03.026

0 引言

在嵌入式系统开发和维护过程中,微控制器 (micro-control unit, MCU) 的软硬件调试是嵌入式系统开发和维护的重要环节,高效强大的调试模块及方法可以缩短系统的开发周期和产品的迭代升级周期,增强产品的竞争力。在航空电子领域,综合化模块化的航电处理平台或设备以其特有的优点被广泛应用。一个机箱内置多个功能集成模块的物理架构是此类航电设备的通用物理架构^[1],即 VPX 机箱技术。VPX 是基于高速串行总线的新一代总线标准,可提供超大

带宽和超大功率,目前被广泛运用于军事等安全性要求较高以及抗恶劣环境下,国产机载计算机设备多遵循 VPX 总线标准设计制造^[2]。VPX 机箱则是基于高速串行总线标准的新一代机箱,其体积坚固,抗干扰能力强且耐震动,内置多个功能模块插槽,普遍应用于服务器、加固计算机等设备,具有高速信号数据采集、实时信号处理及宽频段大容量存储功能^[3]。机载计算机无论是在研发阶段以及在装机后,在返修升级过程中都避免不了调试工作的进行。在这一过程中,对产品进行软硬件调试、功能模块程序加载和软件升级等操作时,针对机箱内部的硬件模块目前主流的方法有两种:一种是直接通过模块的起拔器,将各模块从 VPX 机箱内拔出,通过 JTAG 仿真器连接到 PC 机端进行调试工作,完成

1. 中国航空工业集团公司西安航空计算技术研究所
陕西西安 710068

- [6]VAN H G, COLE E, BEERY S, et al. Benchmarking representation learning for natural world image collections[C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Piscataway: IEEE,2021:12884-12893.
- [7]VAN H G, MAC A O, SONG Y, et al. The inaturalist species classification and detection dataset[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. Piscataway: IEEE, 2018:8769-8778.
- [8]DU R, CHANG D, BHUNIA A K, et al. Fine-grained visual classification via progressive multi-granularity training of jigsaw patches[C]// European Conference on Computer Vision. Cham: Springer International Publishing,2020:153-168.
- [9]YANG X, WANG Y, CHEN K, et al. Fine-grained object clas-

sification via self-supervised pose alignment[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.Piscataway:IEEE,2022:7399-7408.

- [10] LIANG Y, ZHU L, WANG X, et al. Penalizing the hard example but not too much: a strong baseline for fine-grained visual classification[J/OL].IEEE transactions on neural networks and learning systems,2022.[2023-10-16].<https://ieeexplore.ieee.org/abstract/document/9956020>

【作者简介】

黄文龙 (1998—), 男, 安徽淮北人, 硕士, 研究方向: 细粒度分类。

马祥星 (1998—), 女, 安徽淮北人, 本科, 研究方向: 细粒度分类。

(收稿日期: 2023-12-12)